

# The Open Annotation Collaboration: A Data Model to Support Sharing and Interoperability of Scholarly Annotations

**Hunter, Jane**

j.hunter@uq.edu.au  
University of Queensland

**Cole, Tim**

t-cole3@illinois.edu  
University of Illinois, Urbana-Champaign

**Sanderson, Robert**

azaroth42@gmail.com  
Los Alamos National Laboratory

**Van de Sompel, Herbert**

hvdsoomp@gmail.com  
Los Alamos National Laboratory

---

This paper presents the outcomes to date of the annotation interoperability component of the Open Annotation Collaboration (OAC) Project.<sup>1</sup> The OAC project is a collaboration between the University of Illinois, the University of Queensland, Los Alamos National Laboratory Research Library, the George Mason University and the University of Maryland. OAC has received funding from the Andrew W. Mellon Foundation to develop a data model and framework to enable the sharing and interoperability of scholarly annotations across annotation clients, collections, media types, applications and architectures. The OAC approach is based on the assumption that clients publish annotations on the Web and that the target, content and the annotation itself are all URI-addressable Web resources. By basing the OAC model on Semantic Web and Linked Data practices, we hope to provide the optimum approach for the publishing, sharing and interoperability of annotations and annotation applications. In this paper, we describe the principles and components of the OAC data model, together with a number of scholarly use cases that demonstrate and evaluate the capabilities of the model in different scenarios.

## 1. Introduction and Objectives

*Annotating* is both a core and pervasive practice for humanities scholarship. It is used to organize, create and share knowledge. Individual scholars use it when reading, as an aid to memory, to add commentary, and to classify documents. It can facilitate shared editing, scholarly collaboration, and pedagogy. Although there exists a plethora of annotation clients for humanities scholars to use (Hunter 2009) - many of these tools are designed for specific collection types, user requirements, disciplinary application or individual, desktop use. Scholars are also confronted with having to learn different annotation clients for different content repositories, have no easy way to integrate annotations made on different systems or created by colleagues using other tools, and are often limited to simplistic and constrained models of annotations. For example, many existing tools only support the simplistic model in which the annotation content comprises a brief unformatted piece of text. Many tools conflate the storage of the annotations and the target being annotated.

Frameworks for annotation reference are inconsistent, not coordinated, and frequently idiosyncratic, and the constituent elements of annotations are not exposed to the Web as discrete addressable resources, making annotations difficult to discover and re-use. The lack of robust interoperable tools for annotating across heterogeneous repositories of digital content and difficulties sharing or migrating annotation records between users and clients - are hindering the exploitation of digital resources by humanities scholars. Hence the goals of the Open Annotations Collaboration (OAC) are:

- To facilitate the emergence of a Web and Resource-centric interoperable annotation environment that allows leveraging annotations across the boundaries of annotation clients, annotation servers, and content collections. To this end, annotation interoperability specification consisting of an Annotation Data Model will be developed.
- To demonstrate through implementations an interoperable annotation environment enabled by the interoperability specifications in settings characterized by a variety

of annotation client/server environments, content collections, and scholarly use cases.

- To seed widespread adoption by deploying robust, production-quality applications conformant with the interoperable annotation environment in ubiquitous and specialized services and tools used by scholars (e.g., JSTOR, Zotero, and MONK).

In the remainder of this paper we describe related efforts that have informed the development of our Annotation Data Model. We then describe the data model itself that lays a foundation for follow-on work involving demonstrations and reference implementations that exploit real-world repositories such as *JSTOR*, *Flickr Commons*, and *MONK* and leverage existing scholarly annotation applications such *Zotero*, *Pliny* and *Co-Annotea*.

## 2. Related Work

Despite the vast body of work regarding annotation practice, annotation models, and annotation systems, little attention has been paid to interoperable annotation environments. The few efforts in this realm to date comprise:

- RDF-based Annotea developed by Kahan and Koivunen (Kahan et al., 2001);
- Agosti's "A Formal Model of Annotations of Digital Content" (Agosti and Ferro, 2007);
- SANE Scholarly Annotation Exchange;
- OATS (The Open Annotation and Tagging System (Bateman et al., "OATS: The Open Annotation and Tagging System"))).

An analysis of these existing models reveals that on the whole, they have not been designed as Web-centric and resource-centric, or that they have modeling shortcomings that prevent any existing resource from being the content or target of an annotation and from giving an annotation independent status as a resource itself. Further requirements that we have identified that these approaches fail to fully support include:

- Resources of any media type can be Annotation Content or Targets;
- Annotation Targets or Content are frequently segments of Web resources;

- The Content of a single annotation may apply to multiple Targets or multiple annotation Contents may apply to a single Target;
- Annotations can themselves be the Target of further Annotations.

## 3. The OAC Data Model

By exploiting the Web- and Resource-centric approach to modelling annotations, we leverage existing standards and facilitate the interoperability of annotation applications. In the OAC model, an Annotation is an Event initiated at a date/time by an author (human or software agent). Other entities involved in the event are the Content of the Annotation (aka Source) and the Target of the Annotation. The model assumes that the core entities (Annotation, Content and Target) are independent Web resources that are URI-addressable. This approach simplifies and decouples implementation from the repository. An essential aspect of an annotation is the (implicit or explicit) expression of "annotates" relationship between the Content and the Target. The model allows for Content and Target of any media type and the Annotation, Content, and Target can all have different authorship. In situations where the annotation Content or Target is a segment or fragment of a resource (e.g., region of an image), we will draw on the work of the W3C Media Fragments Working Group to specify the fragment address. Figure 1 illustrates the alpha version of the OAC data model.

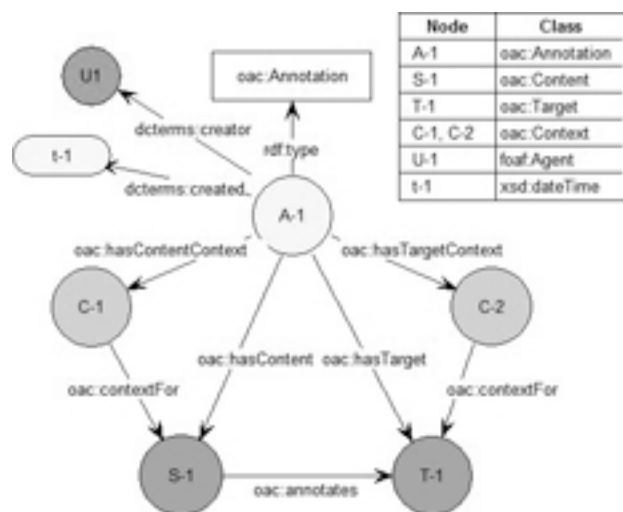


Fig. 1. The Alpha OAC Data Model

## 4. Use Cases

In order to evaluate and demonstrate the feasibility of the OAC Data Model, an initial set of use cases has been developed that are representative of a range of common scholarly practices involving annotation. This preliminary set is available from the OAC Wiki as OAC User Narratives/Use Cases<sup>2</sup> and includes:

- Citation of Non Printed Media
- Commentary on Remote Resources
- Shared Annotations Across Interfaces
- Harvesting, Aggregating, Ranking and Presenting Annotations from Multiple Sites
- Annotating Relationships Between Multiple Mixed-Media Resources
- Annotations which Capture Netchaining Practices
- Annotations with Compound Targets

For example, Figure 2 illustrates a scholarly annotation example involving multiple targets, in which a scholar is making a comment on the differences between segments in scholarly editions of the poem “The Creek of the Four Graves” by Charles Harpur.



Fig. 2. Annotating the differences between two scholarly editions in AustLit

Figure 3 below illustrates the corresponding OAC model for the use case in Figure 2 in which a single annotation Content applies to two Target resources.

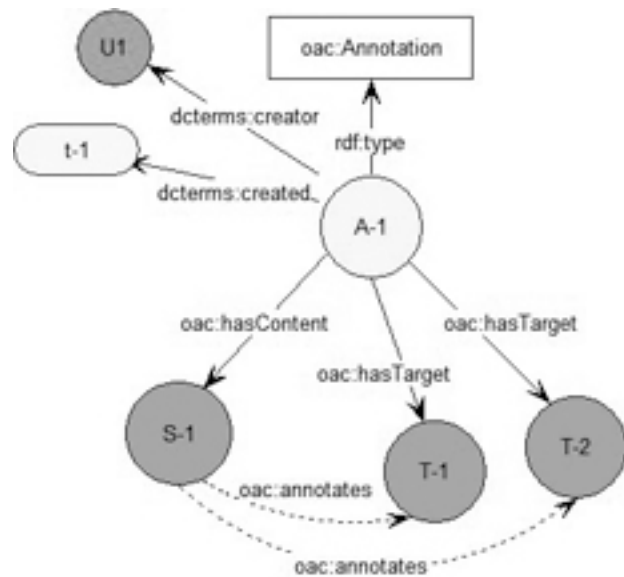


Fig. 3. OAC Model for the example in Figure 2

## 5. Discussion and Conclusions

The proposed OAC Data Model will enable the sharing and discovery of annotations beyond the boundaries of individual solutions or content collections, and hence will allow for the emergence of value-added cross-environment annotation services. It will also facilitate the implementation of advanced end-user annotation services targeted at humanities scholars that are capable of operating across a broad range of both scholarly and general collections. Furthermore, it will enable customization of annotation services for specific scholarly communities, without reducing interoperability. The proposed work will also enable more robust machine-to-machine interactions and automated analysis, aggregation and reasoning over distributed annotations and annotated resources. By grounding our work in a thorough understanding of Web-centric interoperability and embedded models implemented by existing digital annotation tools and services, we create an interoperable annotation environment that will allow scholars and tool-builders to leverage prior tool development work and traditional models of scholarly annotation, while simultaneously enabling the evolution of these models and tools to make the most of the potential offered by the Web environment.

### Acknowledgments

The Open Annotations Collaboration (OAC) is funded by the Andrew W. Mellon Foundation. The authors would also like to acknowledge the

valuable contributions to this work made by: Neil Fraistat, Doug Reside, Daniel Cohen, John Burns, Tom Habing, Clare Llewellyn, Carole Palmer, Allen Renear, Bernhard Haslhofer, Ray Larsen, Cliff Lynch and Michael Nelson. Figure 2 is courtesy of Anna Gerber, Senior Software Engineer on the Aus-e-Lit project.

---

## References

**Hunter J.** (2009). 'Collaborative Semantic Tagging and Annotation Systems'. *Annual Review of Information Science and Technology (ARIST)*. ASIST.

**Kahan, J., Koivunen, M.** (2001). 'Annotea: An Open RDF Infrastructure for Shared Web Annotations'. *Procs of the 10th International conference on the World Wide Web*. Pp. 623-632.

**Agosti, M., Ferro, N.** (2007). 'A Formal Model of Annotations of Digital Content'. *ACM Transactions on Information Systems*. **1**.

*SANE Scholarly Annotation Exchange*. <http://www.huygensinstituut.knaw.nl/projects/sane/>.

**Bateman, S., Farzan, R., Brusilovsky, P., McCalla, G.**. *OATS: The Open Annotation and Tagging System*. <http://fox.usask.ca/files/oats-lornet.pdf>.

**Sanderson R., Van de Sompel H.**. *Open Annotation Collaboration Alpha Data Model Summary*. [http://www.openannotation.org/documents/OAC-Model\\_UseCases-alpha.pdf](http://www.openannotation.org/documents/OAC-Model_UseCases-alpha.pdf).

---

## Notes

1. <http://www.openannotation.org/>
2. <https://apps.lis.illinois.edu/wiki/display/openannotation/OAC+User+Narratives+-+Use+Cases>